ATTORNEY DOCKET NO. 04770.00035 PATENT

Express Mail No. **EV 020 899 115** US

SYSTEM AND METHOD FOR ERROR CONCEALMENT IN DIGITAL AUDIO TRANSMISSION

Title: System and Method for Error Concealment in Digital Audio Transmission

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation-in-part of commonly-assigned U.S. Patent Applications Serial No. 09/770,113 entitled "System and Method for Concealment of Data Loss in Digital Audio Transmission" filed 24 January 2001, and of Serial No. 09/966,482 entitled "System and Method for Compressed Domain Beat Detection in Audio Bitstreams" filed 28 September 2001.

FIELD OF THE INVENTION

[0002] This invention relates to the concealment of transmission errors occurring in digital audio streaming applications and, in particular, to a beat-detection error concealment process.

BACKGROUND OF THE INVENTION

[0003] The transmission of audio signals in compressed digital packet formats, such as MP3, has revolutionized the process of music distribution. Recent developments in this field have made possible the reception of streaming digital audio with handheld network communication devices, for example. However, with the increase in network traffic, there is often a loss of audio packets because of either congestion or excessive delay in the packet network, such as may occur in a best-effort based IP network.

[0004] Under severe conditions, for example, errors resulting from burst packet loss may occur which are beyond the capability of a conventional channel-coding correction method, particularly in wireless networks such as GSM, WCDMA or BLUETOOTH. Under such conditions, sound quality may be improved by the application of an error-concealment algorithm. Error concealment is an important

process used to improve the quality of service (QoS) when a compressed audio bitstream is transmitted over an error-prone channel, such as found in mobile network communications and in digital audio broadcasts.

[0005] Perceptual audio codecs, such as MPEG-1 Layer III Audio Coding (MP3), as specified in the International Standard ISO/IEC 11172-3 entitled "Information technology of moving pictures and associated audio for digital storage media at up to about 1,5 Mbits/s — Part 3: Audio," and MPEG-2 Advanced Audio Coding (AAC), use frame-wise compression of audio signals, the resulting compressed bitstream then being transmitted over the audio packet network. With rapid deployment of audio compression technologies, more and more audio content is stored and transmitted in compressed formats.

[0006] A critical feature of an error concealment method is the detection of beats (i.e., short transient signals) so that replacement information can be provided for missing data. Beat detection or tracking is an important initial step in computer processing of music and is useful in various multimedia applications, such as automatic classification of music, content-based retrieval, and audio track analysis in video. Systems for beat detection or tracking can be classified according to the input data type, that is, systems for musical score information such as MIDI signals, and systems for real-time applications.

[0007] Beat detection, as used herein, refers to the detection of physical beats, that is, acoustic features or other signal transients exhibiting a higher level of energy, or peak, in comparison to the adjacent audio stream. Thus, a 'beat' would include a drum beat, but would not include a perceptual musical beat, perhaps recognizable by a human listener, but which produces little or no sound.

[0008] However, most conventional beat detection or tracking systems function in a pulse-code modulated (PCM) domain. They are computationally intensive and not suitable for use with compressed domain bitstreams such as an MP3

bitstream, which has gained popularity not only in the Internet world, but also in consumer products. A compressed domain application may, for example, perform a real-time task involving beat-pattern based error concealment for streaming music over error-prone channels having burst packet losses.

[0009] The wireless channel is another source of error that can also lead to packet loss. Under such conditions, sound quality may be improved by the application of an error-concealment algorithm. Error concealment is usually a receiver-based error recovery method, which serves as the last resort to mitigate the degradation of audio quality when data packets are lost in audio streaming over error prone channels such as mobile Internet.

[0010] As can be appreciated by one skilled in the relevant art, streaming uncompressed audio over wireless channel is simply an uneconomic use of the scarce resource, and a compressed audio bitstream is more sensitive to channel errors in comparison with an uncompressed bitstream (after removing most of the signal redundancy and irrelevance).

[0011] Conventional error concealment schemes employ small segment (typically around 20 msec) oriented concealment methods including: muting, packet repetition, interpolation, time-scale modification, and regeneration-based schemes. However, a fundamental limitation of packet repetition and other existing error concealment schemes is that they all operate with the assumption that the audio signals are short-term stationary. Thus, if the lost or distorted portion of the audio signal includes a short transient signal, such as a drumbeat, the conventional methods will not be able to produce satisfactory results.

[0012] What is needed is an audio data decoding and error concealment system and method operative in a compressed domain which provides high accuracy with a relatively less complex system at the receiver end.

SUMMARY OF THE INVENTION

[0013] The present invention discloses a beat-pattern based error concealment system and method which detects drum-like beat patterns of music signals on the encoder side of the system and embeds the beat information as data ancillary to a preceding audio data interval in the transmitted compressed bitstream. The embedded information is then used to perform an error concealment task on the decoder side of the system. The beat detector functions as part of an error concealment system in an audio decoding section used in audio information transfer and audio download-streaming system terminal devices such as mobile phones. The disclosed method results from the observation that, while the majority of packet losses in streaming applications are single packet losses, even these single packet losses can result in significant degradation in the subjective audio quality. The disclosed sender-based method improves error concealment performance while reducing decoder complexity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The invention description below refers to the accompanying drawings, of which:

[0015] Fig. 1 is a general block diagram of a conventional audio information transfer and streaming system including mobile telephone terminals;

[0016] Fig. 2 is an illustration of a missing transient signal resulting from conventional error-concealment;

[0017] Fig. 3 is an illustration of a double transient signal resulting from conventional error-concealment;

[0018] Fig. 4 is a general block diagram of a preferred embodiment of a digital audio error concealment system;

[0019] Fig. 5 is a flow diagram illustrating a transmission operation of the error concealment system of Fig. 4;

[0020] Fig. 6 is a flow diagram illustrating a receive operation of the error concealment system of Fig. 4;

[0021] Fig. 7 is a diagram of an encoded bitstream including audio data intervals having short transient signals;

[0022] Fig. 8 is a diagram showing audio data interval updating and replacement via buffers using window type matching;

[0023] Fig. 9 is a flow diagram illustrating the operation of audio data interval updating and replacement in the diagram of Fig. 8;

[0024] Fig. 10 is a diagram of a replacement transient audio data interval disposed between two error-free audio data intervals;

[0025] Fig. 11 is a diagram representing a frequency spectrum of a replacement audio data interval;

[0026] Fig. 12 is a diagram representing a composition operation to form a replacement audio data interval; and

[0027] Fig. 13 is a diagram representing an alternative composition operation to form a replacement audio data interval.

DETAILED DESCRIPTION OF AN ILLUSTRATIVE EMBODIMENT

[0028] Fig. 1 presents an audio information transfer and audio download and/or streaming system 10. System 10 comprises a receiving terminal, such as a mobile phone 11, a base transceiver station 15, a base station controller 17, a mobile switching center 19, a wired telecommunication network 21 such as accessible by a telephone 25, and a telecommunication network 35 accessible by a computer 29 or a user terminal such as a personal digital assistant 27 interconnected either directly or over the computer 29. In addition, there may be provided an audio source, such as a

server unit 31 which includes a central processing unit, memory (not shown), and a database 32, as well as a connection to the telecommunication network 35, which may comprise the Internet, an ISDN network, or any other telecommunication network that is in connection either directly or indirectly to the network into which the mobile phone 11 is capable of being connected, either wirelessly or via a wired line connection. In a typical audio data transfer system, the mobile terminals and the server unit 31 are point-to-point connected.

[0029] Additionally, the telecommunications network 35 and the wired network 21 are interconnected with a wireless telecommunications network 23, which can be a Global System for Mobile Communications (GSM), a General Packet Radio Service (GPRS), Wideband CDMA (WCDMA), DECT, wireless LAN (WLAN), or a Universal Mobile Telecommunications System (UMTS), for example. An alternate audio source can be provided to the wireless telecommunications network 23 via a wireless transceiver 33. Audio signals picked up by a microphone 38 can be encoded by an encoder 37 and provided to the wireless transceiver 33. Alternatively, a source PDA 39 having an internal encoder can provide audio information to the wireless telecommunications network 23 directly through the wireless transceiver 33. Yet another alternative source of audio information is a source mobile phone 13 communicating either directly or indirectly with the base transceiver station 15.

[0030] The user of the mobile phone 11 may select audio data for downloading, such as a short interval of music or a short video with audio music. In a 'select request' from the user, the terminal address of the mobile phone 11 is known to the server unit 31 as well as the detailed information of the requested audio data (or multimedia data) in such detail that the requested information can be downloaded. The server unit 31 then downloads the requested information to another connection end. If connectionless protocols are used between the mobile phone 11 and the server unit 31, the requested information is transferred by using a connectionless connection

in such a way that recipient identification of the mobile phone 11 is thereby connected with the transferred audio information.

[0031] A fundamental shortcoming in the operation of the system 10 can be explained with reference to Fig. 2 in which is shown an audio stream portion 40 such as may be sent to the mobile phone 11 from the server unit 31, from the wireless transceiver 33, or from the source mobile phone 13. The audio stream portion 40 includes an error-free audio data interval (ADI) 41 followed by a defective audio data interval 43. The defective audio data interval 43, which may comprise a corrupted or a missing audio data interval, originally included a short transient signal 45 (where the dashed arrow indicates that the transient signal 45 was corrupted or missing and not received). In a conventional method of error correction, a replacement audio data interval 49 may be substituted for the defective audio data interval 43, as indicated by a replacement arrow 47, to yield an error-concealed audio data stream portion 40'.

[0032] In the example provided, the replacement audio data interval 49 is a copy of the previous error-free audio data interval 41. Because the error-free audio data interval 41 included no transient signal, the replacement audio data interval 49 provides no replacement transient signal for the corrupted or missing short transient signal 45. If the short transient signal 45 comprises a drum beat, for example, the resulting audio stream portion 40' would be conspicuously missing a drumbeat, an effect which would probably be noticed by a user of the mobile phone 11.

[0033] In another application, shown in Fig. 3, an audio stream portion 50 includes an error-free audio data interval 51 followed by a defective audio data interval 53 which originally did not include a short transient signal or drumbeat. In the conventional method of error correction, an error-concealed audio data stream portion 50' is produced by substituting a replacement audio data interval 59 for the defective audio data interval 53, as indicated by a replacement arrow 57. The replacement audio data interval 59 is a copy of the previous error-free audio data

interval 51. However, because the error-free audio data interval 51 included a drumbeat 55, the replacement audio data interval 49 also includes the same drumbeat 55. This conventional error-correction thus produces a double-drumbeat, an effect which would probably be found objectionable by a user of the mobile phone 11. The error-concealment system and method disclosed herein overcomes conventional shortcomings, such as exemplified by the applications of Figs. 2 and 3.

[0034] Fig. 4 presents a generalized block diagram of an error concealment system 60 for digital audio transmission. Operation of the error concealment system 60 can be explained with additional reference to the flow diagrams of Figs. 5 and 6. The error concealment system 60 includes an encoder 61, which may be provided in the server unit 31, the PDA 39, or the source mobile phone 13 (Fig. 1). The error concealment system 60 also includes a decoder 65, which may be provided in the mobile phone 11, the PDA 27, or the computer 29 (Fig. 1). Audio data, such as a musical signal for example, is received at the encoder 61 and may be formatted as a PCM data sample 71, at step 101. The PCM data sample 71 is inputted to the encoder 61 for conversion into audio data intervals, at step 103. The encoder 61 may comprise an encoder based on an MPEG2/4 specification advanced audio encoding (AAC) codec to produce an encoded bitstream 77 such as an MPEG-2 AAC encoded bitstream comprising AAC frames having 1024 frequency components, for example.

[0035] The encoder 61 additionally performs a frequency analysis on the incoming musical signal 71, at step 105, yielding transform coefficients 73 which are used for transient or beat detection. The frequency analysis can use a modified discrete cosine transform (MDCT) to yield MDCT coefficients. In a preferred embodiment, a shifted discrete Fourier transform (SDFT) is used to produce SDFT coefficients. As can be appreciated by one skilled in the relevant art, SDFT is an orthogonal transform and produces more reliable results than MDCT which is not an orthogonal transform. See, for example, the technical paper by Wang, Y., Vilermo, M., and Isherwood, D. *"The Impact of the Relationship Between MDCT and DFT on*

Audio Compression: A Step Towards Solving the Mismatch," ACM Multimedia 2000 International Conference, Oct 30-Nov 4, 2000. The transform coefficients are provided to a transient/beat detector 63 to determine if a current audio data interval includes a transient signal or drumbeat, at decision block 107.

[0036] Preferably, the transient/beat detection is performed using feature vectors (FV), which may take the form of a primitive band energy value, an element-to-mean ratio (EMR) of the band energy, or a differential band energy value. The feature vector can be directly calculated from decoded MDCT coefficients, using the equation for the energy $E_b(n)$ of a band. The energy can be calculated directly by summing the squares of the MDCT coefficients to give:

$$E_b(n) = \sum_{j=N1}^{N2} [X_j(n)]^2 \qquad (6)$$

where $X_j(n)$ is the j[th] normalized MDCT coefficient decoded at an audio data interval n, $N1$ is the lower bound index, and $N2$ is the higher bound index of MDCT coefficients defined in Tables I and II.

Sub-band	Frequency interval (Hz)	Index of MDCT coefficients	Scale factor band index
1	0-459	0-11	0-2
2	460-918	12-23	3-5
3	919-1337	24-35	6-7
4	1338-3404	36-89	8-12
5	3405-7462	90-195	13-16
6	7463-22050	196-575	17-21

Table I. Subband division for long windows

Sub-band	Frequency interval (Hz)	Index of MDCT coefficients	Scale factor band index
1	0-459	0-3	0
2	460-918	4-7	1
3	919-1337	8-11	2
4	1338-3404	12-29	3-5
5	3405-7465	30-65	6-8
6	7463-22050	66-191	9-12

Table II. Subband division for short windows

[0037] If no beat is detected, the current audio data interval can be classified as non-transient and operation proceeds to step 113. If a beat is detected, the current audio data interval is classified as a transient audio data interval, at step 109. The beat information obtained by the beat detector 63 is subsequently embedded within the encoded bitstream 77 as ancillary data or as side information, at step 111, and sent to the decoder 65, at step 113. If there is additional data forthcoming from the server unit 31, at decision block 115, operation returns to step 103. Otherwise, the encoder 61 of the error concealment system 60 stands by for the next audio data request from the mobile phone 11 or other user, at step 117.

[0038] The encoded bitstream 77 is received by a decoder 65, at step 121 in Fig. 6. If the decoder 65 detects no errors in the encoded bitstream 77, at step 123, the audio data intervals comprising the encoded bitstream 77 are converted to a formatted audio sample, such as PCM samples, at step 125. Otherwise, if the decoder 65 detects errors in the received encoded bitstream 77, the corresponding defective audio data interval 81 is provided to an error concealment unit 67. The defective audio data interval 81 is determined as either transient or non-transient, at decision block 127. Ancillary data embedded within the encoded bitstream 77 is used to identify a particular audio data interval as a transient audio data interval 83, as explained in greater detail below.

[0039] Accordingly, a transient defective audio data interval is replaced by an error-free transient audio data interval, at step 129, and converted for output from the

decoder 65, at step 125. Likewise, a non-transient defective audio data interval is replaced by an error-free non-transient audio data interval, at step 131, and converted for output, at step 125. The error concealment unit 67 functions to conceal the detected errors, as described in greater detail below, by returning reconstructed transform coefficients 85, corresponding to the replacement audio data intervals, to the decoder 65 in place of erroneous or missing transform coefficients corresponding to the defective audio data intervals. The decoder 65 utilizes the reconstructed transform coefficients 85 to produce the error-concealed formatted output musical samples 87, at step 125.

[0040] Unlike audio transmission received at the encoder 61, there may be packet loss in the audio transmission transmitted to the decoder 65. This results in certain beats detected by the encoder 61 not reaching the decoder 65. Consequently, beat information obtained by the beat detector 63 at the encoder 61 is more reliable than beat information obtained at the decoder 65. It can thus be appreciated by one skilled in the relevant art that the disclosed error-concealment system and method, which detects beats or transients on the transmitter side, overcomes the limitations of conventional error-concealment systems and methods which perform beat detection on the receiver side.

[0041] There is shown in Fig. 7 an encoded bitstream 150, such as can be transmitted from the encoder 61 to the decoder 65 (Fig. 4). The encoded bitstream 150 includes a transient audio data interval 151 which has a short transient signal 152 here denoted as 'Bassdrum1,' and a transient audio data interval 153 which has a short transient signal 154 here denoted as 'Snaredrum2.' The encoded bitstream 150 also includes a subsequent transient audio data interval 155 with a short transient signal 156 ('Bassdrum3') and a transient audio data interval 157 with a short transient signal 158 ('Snaredrum4'). The signal characteristics of the short transient signals 152 and 156 are similar to one another, and the signal characteristics of the short transient signals 154 and 158 are similar to one another. However, the signal

characteristics of the short transient signals 152 and 156 are different from the signal characteristics of the short transient signals 154 and 158, such as in intensity and/or duration for example, and are accordingly labeled with a different descriptor.

[0042] In a preferred embodiment, the distinction between short transient signals is retained such that if the audio data interval 155 were found to be defective at the decoder 65, the error concealment unit 67 would provide audio data interval 151 as a replacement, as indicated by arrow 169, and not the audio data interval 153. Similarly, if the audio data interval 157 were defective, the audio data interval 153 would be a replacement, as indicated by arrow 183, and not the audio data interval 151. This distinction between two or more different types of transient signals, is provided by a primary set of ancillary beat information 160, or side information, received in the encoded bitstream 150. In the example shown, the ancillary beat information 160 comprises two data bits for each audio data interval in the encoded bitstream 150, including transient audio data intervals 151-157 and audio data intervals 171-177.

[0043] In the diagram, a first data bit 161a ancillary to the audio data interval 171 is used to indicate whether the subsequent audio data interval 151 includes a short transient signal, and a second data bit 161b is used to identify the type of short transient signal present in the subsequent audio data interval 151. The first data bit 161a has a value of '1' to indicate that the audio data interval 151 includes the short transient signal 152, and the second data bit 161b has a value of '1' to indicate that the short transient signal 152 is a 'bassdrum' beat. Similarly, a first data bit 163a ancillary to the audio data interval 173 has a value of '1' to indicate that the subsequent audio data interval 153 includes the short transient signal 154, and the second data bit 163b has a value of '0' to indicate that the short transient signal 154 is a 'snaredrum' beat.

[0044] Thus, if the audio data interval 155 is found to be defective, the error concealment unit 67 reads a first data bit 165a and a second data bit 165b ancillary to the preceding audio data interval 175 to establish that a replacement audio data interval for the defective audio data interval 155 should include a 'bassdrum' short transient signal (i.e., the short transient signal 156). Accordingly, as indicated by the arrow 161, the error concealment unit 67 retrieves the audio data interval 151 from a buffer (such as shown in Fig. 8) as a replacement for the defective audio data interval 155. This method of replacing a defective audio data interval with an error-free audio data interval is referred to in the relevant art as a 'full-band' method of error-concealment.

[0045] Similarly, if the audio data interval 157 is found to be defective, the error concealment unit 67 reads the bits ancillary to the preceding audio data interval 177 to establish that a replacement audio data interval for the defective audio data interval 157 should include a 'snaredrum' short transient signal. The error concealment unit 67 retrieves the audio data interval 153. The error concealment unit 67 uses the replacement audio data interval 153 to reconstruct the transform coefficients 85 associated with the defective audio data interval 157, and sends the reconstructed transform coefficients 85 to the decoder 65 to produce the output musical samples 87.

[0046] It should be understood that that the present invention is not limited to just the one set of ancillary beat information 160 and that a secondary set of ancillary beat information 170 can be used to provide more information in an alternative embodiment and to provide for increased robustness against burst packet loss. In way of example, in the case where both the audio data interval 155 and the preceding audio data interval 175 are lost or corrupted, it is still possible to recover the position of the short transient signal 156 in the audio data interval 155 by obtaining the information provided in additional data bits 167 as indicated by arrow 169. Similarly, for loss of the audio data interval 157 and the preceding audio data interval 177,

recovery is possible by the information provided in additional data bits 181 as indicated by arrow 183.

[0047] In an alternative preferred embodiment, shown in Fig. 8, there is provided in the error concealment unit 67 a first transient buffer 210 storing a plurality of transient audio data intervals 211-217 and a second transient buffer 220 storing a plurality of transient audio data intervals 221-227. Each of the transient audio data intervals 211-217 includes transfer coefficients, such as MDCT coefficients, for a first type of short transient signal or beat, each beat here denoted as a 'TransientA' type of beat (as represented by a triangular arrowhead), and each of the audio data intervals 221-227 includes transfer coefficients for a second type of short transient signal or beat, here denoted as a 'TransientB' type of beat (as represented by a round arrowhead). TransientA can represent a bassdrum beat, and TransientB can represent a snaredrum beat in accordance with the examples provided above.

[0048] As understood by one skilled in the relevant art, MP3 applications, for example, use four different window types for sampling: a long window, a long-to-short window (i.e., a 'stop' window), a short window, and a short-to-long window (i.e., a 'start' window). These window types are indexed as 0, 1, 2, and 3 respectively. Accordingly, each of the transient audio data intervals 211-217 comprises the same type of beat but a different window type. For example, the audio data interval 211 includes a TransientA type of beat in a type-0 window, the audio data interval 213 includes a TransientA type of beat in a type-1 window, and so on as indicated by the subscripts. Similarly, each of the audio data intervals 221-227 includes a TransientB type of beat with a different window type, as indicated by subscripts.

[0049] The functions performed using the transient buffers 210 and 220 can be described with additional reference to the flow diagram of Fig. 9. The decoder 65 (Fig. 4) operates to decode audio data intervals received in the encoded bitstream 77,

a portion of which is represented by a disjoint series of audio data intervals 200-207 on a time coordinate 209 in Fig. 8. The decoder 65 decodes the next audio data interval in the encoded bitstream 77, at step 281, represented here by an audio data interval 200. The decoder 65 checks the audio data interval 200 for ancillary data pertaining to beat information in the next audio data interval 201. If there is no ancillary data provided, operation returns to step 281. If, at decision block 283, ancillary transient data 200a is present, the bits '1' and '1' are used to determine that, if error-free, the next audio data interval 201 includes a TransientA beat, at step 285. The next audio data interval 201 is decoded, at step 287, and a query is made as to whether the audio data interval 201 is defective, at decision block 289.

[0050] If the audio data interval 201 is error-free, the TransientA buffer 210 is updated with the audio data interval 201, as indicated by arrow 231. In the example provided, the audio data interval 201 includes a beat in a type-2 window. Accordingly, transform coefficients in the buffered transient audio data interval 215 are replaced by the transform coefficients in the decoded audio data interval 201, at step 291, and operation returns to step 281. At some later time, the decoder 65 determines from an audio data interval 202 that the next audio data interval 203 should be a transient audio data interval with a TransientB-type beat. Accordingly, if the transient audio data interval 203 is error-free, the second transient buffer 220 is updated by replacing the buffered type-0 window transient audio data interval 221 with the decoded transient audio data interval 203, as indicated by arrow 233.

[0051] If, at decision block 289, a transient audio data interval is found to be defective, the decoder goes to a buffer corresponding to the transient type and to the window-type missing from the defective transient audio data interval, at step 293, and the correct transient audio data interval is retrieved from the correct transient buffer for replacement, at step 295. The retrieved transient audio data interval is substituted for the defective transient audio data interval, at step 297, and operation returns to step 281. In the example provided, an audio data interval 205 is found to be

defective. From the preceding transient audio data interval 204, which is a type-2 window and which includes the bits '1' and '1' in the ancillary data, the decoder 65 determines that the defective transient audio data interval 205 originally included a TransientA-type beat in a type-3 window. This determination is made on the expected occurrence of a type-3 window following a type-2 window in the proximity of a transient. Accordingly, the defective transient audio data interval 205 is replaced by transient audio data interval 217 obtained from the first transient buffer 210. Likewise, for a defective transient audio data interval 207, information obtained from a preceding audio data interval 206 indicates that the original transient audio data interval 207 included a TransientB-type beat in a type-1 window. Accordingly, a transient audio data interval 223 is selected for replacement of the defective transient audio data interval 207.

[0052] There is shown in Fig. 10, a diagrammatical illustration of an encoded bitstream segment 240 including an error-free (n-1)th audio data interval 241 and an error-free (n+1)th audio data interval 243. An nth audio data interval (not shown) originally transmitted between the (n-1)th audio data interval 241 and the (n+1)th audio data interval 243 was found to be defective and, accordingly, was replaced by a replacement audio data interval 245 comprising a drumbeat 247 and harmonic structure 249 adjacent the drumbeat 247. The harmonic structure 249 is provided by copying from a previous audio data interval (not shown) associated with the replacement drumbeat 247. Accordingly, there results a discontinuity in the harmonic structure from the audio data interval 241 to the harmonic structure 249, and from the harmonic structure 249 to audio data interval 243. This audio discontinuity has been referred to in the relevant art as a 'spectral fine structure disruption effect.'

[0053] To mitigate this effect, a sub-band method of audio data interval replacement can be used in place of the full-band method described above. The sub-band method can be explained with reference to the diagram in Fig. 11 in which is shown an audio data interval frequency band 250 divided into a low-frequency band

251 (i.e., frequency range F_0 to F_1), a mid-frequency band 253 (i.e., frequency range F_1 to F_2), and a high-frequency band 255 (i.e., frequency range F_2 to F_3). The mid-frequency band 253 represents the most relevant harmonic and melodic parts of the audio data signal. The low-frequency band 251 and the high-frequency band 255 are more relevant for the drumbeat. In an alternative preferred embodiment, the low-frequency band 251 and the high-frequency band 255 are copied from a previous beat containing an appropriate drum beat (not shown), and the mid-frequency band 253 is copied from a neighboring audio data interval, for example from the audio data interval 241 (Fig. 10) for replacement as the harmonic structure 249. In one preferred embodiment, F_1 is approximately 344 Hz, and F_2 is about 4500 Hz. These values were obtained empirically based on the spectrogram observation of relevant test signals and the constraints of the AAC standard. In way of example, F_1 corresponds to the 16^{th} MDCT coefficient for a long type-0 window, and F_2 corresponds to the 208^{th} MDCT coefficient. For a short type-2 window, F_1 corresponds to the 2^{nd} MDCT coefficient, and F_2 corresponds to the 26^{th} MDCT coefficient.

[0054] This method is shown in greater detail in Fig. 12 as a composition or mixing operation used to produce a replacement audio data interval 265. This composition method combines a first audio data interval 261, denoted by $X(r)$, and a second audio data interval 263, denoted by $Y(r)$ to produce a composite audio data interval, denoted by $Z(r)$. The first audio data interval 261 comprises the spectral data from a previous beat or transient signal, such as may be obtained from a transient buffer. The second audio data interval 263 comprises an audio data interval (not shown) in a transfer domain preceding the defective audio data interval. The replacement transfer coefficients for the defective audio data interval are given by $Z(r)$:

$$Z(r) = \alpha(r)X(r) + \beta(r)Y(r), \quad 0 \le r \le N-1 \qquad (1)$$

where $\alpha(r)$ and $\beta(r)$ are weighting functions across the entire frequency band with constraints of

$$\alpha(r)+\beta(r)=1, \quad 0 \le r \le N-1 \tag{2}$$

and

$$\alpha(r),\beta(r) \ge 0, \quad 0 \le r \le N-1 \tag{3}$$

[0055] The parameters $\alpha(r)$and $\beta(r)$ can be adaptive to the actual signal, or can be static parameters for simplicity. The design principle is to maintain the harmonic continuity while keeping the beat structure in place. A simple implementation can be

$$\alpha(r)=\begin{cases}0, & F_1 < r \le F_2 \\ 1, & elsewhere\end{cases} \tag{4}$$

$$\beta(r)=\begin{cases}1, & F_1 < r \le F_2 \\ 0, & elsewhere\end{cases} \tag{5}$$

where $z(k)$ is an output audio signal 267 after application of an inverse transform, such as an inverse modified discrete cosine transform (IMDCT), of $Z(r)$:

$$z(k) = IMDCT(Z(r)) \tag{6}$$

[0056] The audio data interval 265 formed by the function $z(k)$ is used as a replacement for the defective audio data interval. This method has low computational complexity and low memory requirements in the decoder 65 and can be advantageously used in smaller devices such as the mobile phone 11.

[0057] For better performance, an alternative embodiment of the disclosed method is illustrated in Figure 12. The two signals, $x(k)$ and $y(k)$, are first weighted in the frequency domain before inversely transforming back to time domain. For MDCT transform,

$$x(k) = IMDCT[\alpha(r)X(r)] \tag{7}$$

$$y(k) = IMDCT[\beta(r)Y(r)] \tag{8}$$

where $\alpha(r)$ and $\beta(r)$ are weighting functions in the frequency domain similar to the weighting functions in equation (1). The replacement signal $z(k)$ is then constructed as

$$z(k) = a(k)x(k) + b(k)y(k), \quad 0 \le k \le 2N-1 \tag{9}$$

where $a(k)$ and $b(k)$ are weighting functions in the time domain with constraints of

$$a(k) + b(k) = 1, \quad 0 \le k \le 2N-1 \tag{10}$$

$$a(k), b(k) \ge 0, \quad 0 \le k \le 2N-1 \tag{11}$$

[0058] The parameters $a(k)$ and $b(k)$ can be adaptive to the actual signal or static. The design principle is to estimate the drum contour in time domain. For a simple implementation, $a(k)$ can be a static function such as a triangle function 271 to approximate the drum contour in time domain. The asymmetric triangle 273 indicates that the onset of a drum is generally much shorter than the subsequent decay. The term T_B indicates the maximum of the weighting function $a(k)$.

[0059] The above is a description of the realization of the invention and its embodiments utilizing examples. It should be self-evident to a person skilled in the relevant art that the invention is not limited to the details of the above presented examples, and that the invention can also be realized in other embodiments without deviating from the characteristics of the invention. Thus, the possibilities to realize and use the invention are limited only by the claims, and by the equivalent embodiments which are included in the scope of the invention.

[0060] What is claimed is: